Non-pest household arthropods as a reservoir of human opportunistic pathogens

Federica Boiocchi[1], Romaine Derelle[2], Matthew Davies[3], Luisa Orsini[2], and Anthony Hilton[1]

[1]Aston University
[2]University of Birmingham
[3]Killgerm Chemicals Ltd

December 5, 2021

Abstract

Arthropods are recognised as potential mechanical and biological vectors for infectious diseases in outdoor environments. However, a comprehensive understanding of the indoor arthropod community diversity and of the role that their associated microbiota may have as disease vectors is largely unexplored. Here, we study the arthropod community and the associated microbiota diversity of twenty indoor environments, sampled over a period of twelve months from urban and suburban households by citizen scientists in the West Midlands (UK). We compare the arthropods diversity between environments and over the sampling months. We characterize the exogenous (exoskeleton) and endogenous (gut) bacterial communities associated with all specimens of arthropods actively captured using both a traditional culture-based and an unbiased metabarcoding approach. For the first time, we describe the exogenous and endogenous microbiota composition and diversity of 14 arthropod families found in indoor environments. We find that both the exogenous and the endogenous microbiota are potential carriers of human opportunistic pathogens, with potential implications for public health. We discover that many bacteria families are shared across the exogenous microbiota of arthropods, likely influenced by the bacteria present in the environment. Conversely, the endogenous microbiota composition is unique to the arthropod families, and likely genetically determined. We show that the metabarcoding unbiased approach is a superior tool to characterize the microbiota associated with each arthropod family. This study provides new insights into bacterial carriage in household arthropods as potential reservoirs of infectious disease.

Introduction

As common organisms in indoor environments, arthropods are recognised mechanical and biological vectors of infections in animals, including humans (Doron & Gorbach 2008). Arthropod pests are mechanical carriers of human pathogens, especially of gastrointestinal and other opportunistic pathogens, including *Escherichia* ,*Salmonella* , *Shigella* and *Campylobacter* (Adams*et al.* 2015; Gibbons 2016). *Klebsiella pneumoniae*and *Staphylococcus aureus* have been found to be common in the microbiome of cockroaches collected in hospitals, whereas the same arthropod species carries more frequently *Escherichia coli, Citrobacter* spp., *Pseudomonas aeruginosa* and*Staphylococcus* spp. in domestic households (Memona *et al.*2017; Menasria *et al.* 2015; Moges *et al.* 2016). Flies are also known carriers of opportunistic bacteria, such as *Klebsiella pneumoniae* (Ranjbar *et al.* 2016), *Staphylococcus* and Enterobacteriaceae (Barreiro *et al.* 2013).

Gastrointestinal pathogens carried by flies and cockroaches can cause 2.4 million infections and 180 deaths per year in the UK (Holland*et al.* 2020) . Worldwide, pest arthropods are one of the main factors involved in the epidemiology of diarrhoea (Rahmadani & Lee 2020), responsible for the yearly death of 525,000 children under the age of five (Das *et al.* 2018). Pest arthropods play such an important role in opportunistic infections that the control of e.g. fly populations can reduce the incidence of gastrointestinal bacterial in-

fections by 23% (Das *et al.* 2018; Knight *et al.* 1992; Sengupta *et al.* 1995). However, arthropod pests comprise a minority of the indoor arthropod environment (Barberan *et al.*2015; Bertone *et al.* 2016; Leong *et al.* 2017); studies start to emerge describing the abundance and diversity of non-pest arthropods in these environments (Adams *et al.* 2015; Barberan*et al.* 2015; Bertone *et al.* 2016; Leong *et al.*2017). The role of these species as potential carriers of opportunistic infections is, to date, understudied. However, human pathogenic bacteria have been isolated from the microbiome of non-pest arthropods, including dark-winged fungus gnats (Diptera: *Sciaridae*) (Kwon *et al.*2016); owl midges (Diptera: *Psychodidae*) (Faulde & Spiesberger 2013) , and cellar spiders (Araneae: Pholcidae) (Voloshyn *et al.*2017). In addition, studies on household surfaces have shown the presence of both human-associated pathogens, including*Staphylococcus* , *Streptococcus, Corynebacterium* ,*Lactococcus* and *Enterobacteriaceae* , as well as bacteria associated with the outdoor environment, such as *Pseudomonas* ,*Propionibacterium* , *Streptophyta* and *Sphingomonas*(Jeon *et al.* 2013). However, a systematic characterization of non-pest arthropods diversity and their associated microbiota as potential carriers of human pathogens in indoor environments is lacking. Moreover, the screening of pathogenic bacteria associated with arthropods is conducted prevalently with culture-based approaches. Culturable bacteria only constitute 2.5% of existing bacteria, severely limiting the characterization of bacterial communities associated with the indoor environment (Oberauner *et al.* 2013) .

Here, we study the indoor arthropod community and its associated microbiota, sampled from twenty households located in the West Midlands (UK). The households were evenly split between urban and suburban environments and sampling was conducted by citizen scientists through active captures over a period of 12 months. We classified the arthropods and their relative abundance across the 12-month sampling period using taxonomic cues against entomological references. We applied DNA metabarcoding and culture-based approaches to study the composition and relative abundance of endogenous (gut) and exogenous (exoskeleton) microbiota of the captured arthropod morphospecies. We determined whether the microbiota dynamics and composition were explained by the association with the arthropod morphospecies, the sampling environment (urban/suburban) and the season (month of collection). We then determined how many bacteria families identified in the arthropod microbiota were recognized as human pathogens according to the list of biological agents by the Advisory Committee on Dangerous Pathogens (HSE). Our study provides important insights into the role of non-pest arthropods as potential carriers of opportunistic pathogens for animals and humans, with implications for public health.

Materials and Methods

Sampling and experimental design

Arthropods were collected monthly from 20 households between August 2018 and October 2019 by citizen scientists in the West Midlands, UK (Fig. 1). Nine households were in sub-urban and 11 were in urban settings; the active collection was done using a spider catcher (Brainstorm, Gisburn, UK), and a pooter bug catcher (Ocean, Southampton, UK) (Fig. 1). Immediately after collection, the arthropods were stored at -20 in domestic freezers until monthly collection by the laboratory staff at Aston University (UK), where samples were taxonomically identified at the family level against entomological references (Chinery 1993; Chinery & Falk 2007) using a stereo microscope (Motic Hong Kong Limited, Hong Kong, China).

Microbiome characterization

We used both a culture-based microbial analysis followed by Sanger sequencing and DNA metabarcoding to quantify the microbial communities associated with household arthropods captured by the citizen scientists (Fig. 1). For samples collected between August 2018 and July 2019, we characterized the composite microbiota associated with the arthropods' morphospecies, including exogenous (exoskeleton) and the endogenous (gut) microbiota together. For a subset of the samples (July to October 2019), we analysed the exogenous and the endogenous microbiota separately. This allowed us to study whether the exogenous microbiota composition varied seasonally and co-varied with environmental variables (e.g., temperature and humidity).

The endogenous and exogenous microbiota were collected from pools of up to 5 arthropods from the same morphospecies to capture the bacterial communities' diversity. The exogenous microbiota was collected

by washing and vortexing the arthropod pools (2850 rpm for 30 seconds) in sterile distilled water (SDW) and pelleting the bacterial community through centrifugation at 13000g for two minutes in an Eppendorf centrifuge (Hamburg, Germany). Following the collection of the exogenous microbiota, the pools were washed in 70% ethanol for 2 minutes, rinsed with SDW, and homogenised by grinding in 1 ml of SDW using a sterile micro pestle (Eppendorf, Hamburg, Germany), to collect the endogenous (gut) microbiota. Both the total microbiota extracted from samples collected between August 2018 and July 2019 and the microbiota collected separately from endogenous and exogenous microbiota (July to October 2019), were split into two aliquots; the first aliquot was serially diluted down to 10^{-4} of the initial concentration and plated on Nutrient Agar, Mannitol Salt Agar, Violet Red Bile Glucose Agar and Blood Agar N° 2 (Thermo Fisher Scientific Oxoid Ltd, Basingstoke, UK); the second aliquot was used for DNA metabarcoding (Fig. 1).

All the inoculated bacteria on agar media were incubated at 37°C for 48h in aerobic conditions, whereas Blood agar plates were incubated both in aerobic and anaerobic conditions. Following incubation, bacterial colonies were counted, corrected for the number of arthropods in each pool and the serial dilution factor and expressed as the number of Colony Forming Units (CFU/ml/arthropod). A total of 166 colonies across the collected samples was sequenced using Sanger sequencing technology, following gDNA extraction with QIAamp DNA mini kit (Qiagen, Hilden, Germany), and amplification of the 16S rRNA Vi-V2 region [27F (5'-AGAGTTTGATCATGGCTCA-3') and 1492R (5'-TACGGTTACCTTGTTACGACTT-3')]. The PCR conditions were as follows: 95°C for 15 minutes, followed by 35 cycles at 95°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute. A final extension at 72°C for 10 minutes was used. The PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and quantified on a 1.2% agarose gel stained with ethidium bromide (Bio-Rad Laboratories, Hercules, California, USA). The Sanger sequencing was completed by Eurofins Genomics, Ebersberg, Germany.

Genomic DNA for the metabarcoding analysis was extracted from the microbiota community of each arthropod pool using the DNeasy PowerWater kit (Qiagen, Venlo, Netherlands) following filtration on 0.2 μm filter funnels (Thermo Fisher Scientific, Waltham, USA). Extracted DNA was then prepared in paired end 250bp amplicon libraries obtained with the 16S RNA V1-V2 region, using a 2 step PCR protocol with 96x96 dual tag barcoding to facilitate multiplexing and to reduce crosstalk between samples in downstream analyses (MacConaill *et al.* 2018). Negative controls were used for PCR biases and contamination. They consisted of genomic libraries constructed using the SDW used to wash the arthropods and of libraries without target DNA. PCR1 and PCR2 primers, as well as annealing temperatures per primer pair in PCR1 are in Table S1. QCs were performed at each PCR step. Excess primer dimers and dinucleotides from PCR1 were removed using Thermostable alkaline phosphatase (Promega) and Exonuclease I (New England Biolabs). PCR2 amplicons were purified using High Prep PCR magnetic beads (Auto Q Biosciences) and quantitated using a 200 pro plate reader (TECAN) using qubit dsDNA HS solution (Invitrogen). A standard curve was created by running standards of known concentration on each plate against which sample concentration was determined. PCR2 amplicons were mixed in equimolar quantities at a final concentration of 12 pmol using a Biomek FXp liquid handling robot (Beckman Coulter). Pooled libraries molarity was confirmed using a HS D1000 tapestation screentape (Agilent) prior to 250bp paired end sequencing on an Illumina MiSeq platform to obtain 100,000 reads per sample and amplicon. The genomic libraries were prepared and sequenced by EnviSion, Environmental BioSequencign and BioComputing at the University of Birmingham (*https://www.envision-service.com/*).

Data analysis

The sequences obtained with Sanger technology were analysed using the software FinchTv (version 1.4.0) and taxonomically assigned using Blast V 2.12.0+ using default parameters, following the removal of PCR primer sequences.

Metabarcoding sequences were analysed using qiime2 (Caporaso *et al.* 2010). Amplicon sequence variants (ASVs) were obtained by removing adapter primers with the cutadapt plug-in (Martin 2011), denoising with DADA2, trimming low quality reads, merging forward and reverse reads, dereplicating and filtering out

chimaeras (Callahan *et al.* 2016). Taxonomic assignment of the 16S sequences was done using the DAIRY database (Meola*et al.* 2019) , following the removal of unassigned reads with the microDecon R package (McKnight *et al.* 2019).

The R-package Vegan (Oksanen 2020) was used for statistical analysis of both the culture-based and the metabarcoding sequences. We applied a permutation analysis of variance (PERMANOVA) on the weighted Bray-Curtis distance for both metabarcoding and culture-based sequences, using the Adonis function in R (Team 2020) to quantify the effect of Arthropod family, season (month of sampling) and environment (urban/suburban), and their combined effects on the microbiota diversity.

We studied correlations between environmental variables and the compositional changes of the exogenous and endogenous microbiota. The environmental variables studied were average air temperature, precipitation, humidity, and wind speed obtained from the Stourbridge weather station (UK), located within a 20 Km range from the 20 households used in this study (metnet.co.uk). Significant associations between beta diversity (weighted diversity between samples) and environmental variables were established using a Spearman's rank correlation test ($P < 0.05$).

The bacterial families identified through the culture-based and high throughput DNA metabarcoding were mapped onto the approved list of biological agents published by the Health and Safety Executive Advisory Committee on Dangerous Pathogens ((ACDP) 2020) to identify pathogens with potential adverse effect on human health.

Results

Arthropod captures across seasons and households

Over the 12 months collection period we identified 215 arthropods using taxonomic keys. *Pholcidae* (cellar spiders), were the most abundant family, accounting for 33.5% of the total captures, closely followed by *Coccinellidae* (ladybirds; 32.1%) and*Lepismatidae* (silverfish; 12.1%). Rarer specimens included*Chironomidae* (non-biting midges; 6%), *Calliphoridae* (blue bottle flies; 3.3%), *Gnaphosidae* (ground spiders; 2.8%),*Culicidae* (mosquitoes; 2.3%), *Pyralidae* (pyralid moths; 1.9%), *Tipulidae* (crane flies; 1.4%), *Agelenidae* (funnel weavers; 0.9%), *Blattidae* (cockroaches; 0.9%), *Noctuidae*(noctuid moths; 0.9%), *Thomisidae* (crab spiders; 0.9%) and*Vespidae* (wasps; 0.9%) (Fig. 2).

Household arthropod- associated microbiota

We studied the microbiota composition associated with the arthropods captures and measured how they varied by arthropod's family, season (month of sampling) and environments (urban/suburban).

The microbiota identified with the culture-based approach significantly differed among the arthropod families, and between environments, whereas it did not vary by season (Table 1A; Fig. 3A). We cultured and isolated bacteria from 84 pools (72.6%), whereas the remaining 31 pools (27.4%) did not harbour any culturable bacteria. Among the bacteria that could be cultured and taxonomically identified, Gram-positive cocci (including those strains not identifiable at family level) and Gram-positive rods were the most abundant (Fig. 3A, Table S2), followed by*Staphylococcaceae* (30.36%), *Bacillaceae* (10.71%) and*Micrococcaceae* (10.12%) (Fig. 3A, Table S2).*Streptococcaceae* and *Enterococcaceae* accounted for 9.52% and 8.33% of the taxonomically assigned sequences, respectively (Fig. 3A; Table S2). The remainder of the taxonomically assigned reads belonged to: *Moraxellaceae* (4.76%), *Dysgonomonandaceae*(4.17%), *Rhodobacteraceae* (4.17%), *Yersiniaceae*(4.17%), *Erwiniaceae* (2.98%), *Carnobacteriaceae*(1.79%), *Clostridiaceae* (1.79%), *Brevibacteraceae*(1.19%) and *Microbacteriaceae* (1.19%) (Fig. 3A, Table S2). Rare families accounting for 0.6% each of the total identified sequences were: *Peptostreptococcaceae* , *Caryophanaceae* ,*Corynebacteriaceae* , *Dermabacteraceae* ,*Enterobacteriaceae* , *Neisseriaceae* ,*Pseudomonandaceae* and *Veillonellaceae* (Table S2). A total of 192 isolates could not be identified at family level; they were Gram-positive cocci (49.48% of the unassigned sequences), Gram-positive rods (27.6%), Gram-negative rods (15.10%), Gram-negative cocci (6.77%), Gram-negative coccobacilli (0.52%) and Gram-negative filamentous bacteria (0.52%) (Table S2).

Gram-positive cocci and Gram-positive rods were found in all arthropods, except for Vespidae (Gram-positive cocci) and Blattidae (Gram-positive rods) (Fig. S1A). Even if few bacteria families were shared across the artropod's families, each arthropod had a unique microbiota composition (Fig. S1A).

The microbiota identified with the metabarcoding approach significantly differed among arthropod families, season (sampling months) and environments (urban/suburban) (Fig. 3B; Table 1). The interaction term Arthropod Family*Season was also significant (Table 1). The metabarcoding approach identified 771 ASVs, taxonomically assigned to 15 phyla, 111 families and 249 genera of bacteria. Among the 10 most abundant families, *Rickettsiaceae* comprised 67.3 % of the identified families, followed by *Anaplasmataceae* (10.59%), unidentified family of the class Bacilli (3.59%),*Proprionibacteriaceae* (3.71%), *Porphyromonadaceae*(3.29%), *Micrococcaceae* (2.89%), *Enterobacteriaceae*(2.89%), *Enterococcaceae* (2.12%), *Methylobacteriaceae*(1.83%) and an unidentified family of the order *Burkholderiales*(1.78%) (Fig. 3B; Table S2). The metabarcoding approach revealed that*Rickettsiaceae* were common and abundant in the microbiota of all arthropod families, followed by *Anaplasmataceae* (Fig. S1B; Table S2). *Bacilli* were abundant in *Tipulidae* ; they were present in low abundance in 50% of the captured arthropods (Fig. S1B). Although the microbiota of the different arthropod families was dominated by*Rickettsiaceae* , the relative abundance of other bacterial families was distinct among the arthropods (Fig. S1B, Table S2). For example, *Enterobacteriaceae* and *Micrococcaceae* made up 15.2 and 13.9%, respectively, of the Agelenidae microbiota, whereas*Enterococcaceae* (14.4%) were the second most abundant bacteria family in *Blattidae* (Fig. S1B; Table S2). The microbiota of*Calliphoridae* was dominated by *Porphyromonandaceae*(70.7%), whereas *Rickettsiaceae* (8.3%) accounted for a small proportion of this microbiota (Fig. S1B).

We studied the change in community composition of the exogenous and endogenous microbiota separately in the arthropod captures between July and October 2019. We asked if changes in the endogenous and exogenous microbiota were explained by the arthropod family, the month of collection and their interaction terms. We also studied the correlation between the exogenous and endogenous microbiota communities with environmental variables. In the culture-based approach, the endogenous and exogenous microbial communities varied significantly by the arthropod families, whereas they did not differ significantly among seasons (month of collection; Fig. 4A; Table 2). Gram-positive cocci were abundant in both the endogenous and exogenous microbiota (Fig. 4A). Some bacteria families were only found in the exogenous microbiota (*Staphylococcaceae* , *Peptostreptococcaceae* ,*Bacillaceae* , *Enterococcaceae* , *Rhodobacteraceae* and*Erwiniaceae* were isolated from the exoskeleton of the arthropods), whereas others were unique to the endogenous microbiota (*Enterococcaceae* , *Erwiniaeae* , *Rhodobacteraceae* ,*Micrococcaceae* , *Carnobacteriaceae* and*Streptococcaceae*) (Fig. 4A). Some bacterial families were arthropod-specific: *Staphylococcaceae* were only found in the exogenous microbiota of *Coccinellidae* and *Blattidae* ;*Rhodobacteraceae* were only found in the endogenous microbiota of Blattidae and the exogenous microbiota of Gnaphosidae (Fig. 4A).*Carnobacteriaceae* and *Streptococcaceae* were only found in the *Coccinellidae* 's endogenous microbiota (Fig. 4A).

The endogenous and exogenous microbiota identified with the metabarcoding approach were overall divergent and varied significantly by the arthropod family, whereas they did not differ significantly by the sampling month (Table 2). *Rickettsiaceae* were both common and abundant across the arthropods' families (Fig. 4B). The exogenous microbiota was dominated by Rickettsiaceae, comprising 64.1% of the microbiota in *Blattidae* , 30.2% in *Coccinellidae* , 55.8% in Gnaphosidae spiders, 76.3% in *Pholcidae* and 34.4% of in*Tipulidae* . The exogenous microbiota of Coccinellidae and*Tipulidae* were the most diverse among the arthropods studied here. In addition to *Rickettsiaceae* , *Coccinellidae*exogenous microbiota included *Propionibacteriaceae* (18.1%),*Methylobacteriaceae* (13.7%), *Micrococcaceae* (11.4%), and*Streptococcaceae* (9.4%) in relatively high abundance (Fig. 4B). The exogenous microbiota of *Tipulidae* comprised*Enterobacteriaceae* (23.7%), Bacilli (16.9%),*Planococcaceae* (6.1%), *Streptococcaceae* (5%), in addition to *Rickettsiaceae* (Fig. 4B). The endogenous microbiota composition was arthropod-specific. For example, *Enterococcaceae*made up 94% of the endogenous microbiota of *Blattidae* , whereas they were in low abundance (<1%) in other arthropods (Fig. 4B). *Bacilli* comprised 93% of the endogenous microbiota in*Tipulidae* and 42% in *Coccinellidae* (Fig. 4B).*Anaplasmataceae* were present in low abundance in the endogenous microbiota of all arthropods and abundant (35%) in *Gnaphosidae*(Fig. 4B).

We used the Spearman correlation combined with the Shannon-Wiener index, which combines species richness and relative abundances, to establish associations between endogenous/exogenous bacterial communities and environmental variables: temperature, precipitation, humidity, and wind speed. Neither the endogenous nor the exogenous microbiota compositional changes were explained by significant covariation with the environmental variables (Table S3).

Pathogenic bacteria associated with household arthropods

The bacteria identified through the culture-based and the metabarcoding approach were intersected with the Health and Safety Executive (HSE) list of biological agents, in order to assess the presence of pathogenic bacteria associated with the household arthropods. The metabarcoding and the culture-based approach shared 10 (4.7%) bacteria families, which were all listed in the HSE database as potentially pathogenic (Fig. 5). A single bacterial family that intersected with the HSE database was identified solely through the culture-based approach, whereas 18 families (8.5%) uniquely identified with the metabarcoding approach intersected with the HSE database (Fig. 5). A significant positive correlation based on the 2x2 contingency table was found between isolated human pathogens and indoor arthropods, as well as between environmental bacteria and outdoor arthropods ($\chi 2$=28.177, p = 0.0001) (Table S4).

Discussion

Non-pest arthropods are a potential reservoir of human pathogens

Most studies of indoor environments have focused on pest arthropods, neglecting non-pest species and leading to an underestimation of the potential impact of non-pest arthropods as carriers of human pathogens (Bertone *et al.* 2016). Here, we characterized the indoor insect community and the associated microbiota over a period of 12-month across 20 households. This approach enabled us to characterize for the first time the microbiota of the arthropod indoor community and to identify potential carriers of human pathogens.

The most abundant arthropod family in the indoor environments studied here were cellar spiders of the family Pholcidae, consistent with previous findings in indoor environments (Desales-Lara *et al.*2013; Durán-Barrón *et al.* 2009; Rodríguez-Rodríguez *et al.*2015). Cellar spiders, together with ladybirds and cockroaches are typically found in indoor environments (Brown *et al.* 2008; Wang*et al.* 2011). However, the other arthropods found in our study were unexpected as they are outdoor species (e.g. wasps and crane flies) (Hall & Gerhardt 2002; Reed & Landolt 2019). They were likely casual intruders attracted by food and/or light (e.g., non-biting midges and spiders). These species are potential carriers of human pathogens, as their microbiota composition showed.

Our survey of the microbiota associated with non-pest arthropods identified opportunistic pathogens of humans present either in the endogenous or exogenous microbial community or both. Among the bacterial families intersecting with the HSE database, Rickettsiaceae were ubiquitous to the endogenous and exogenous microbiota of all indoor arthropods actively captured in this study. *Rickettsia* is a genus of Gram-negative obligate intracellular bacteria typically associated with ticks, leeches, amoeba, ciliate, and hydra (Krawczak *et al.*2018). They can cause typhus and spotted fever, with potential life-changing illness (Ogata *et al.* 2006). *Bacillaceae*were the second most abundant bacteria family across the insect captures, which is responsible for foodborne infections in humans (Stenfors Arnesen *et al.* 2008). The genus *Clostridiaceae*was also common across the insect captures; it comprises several spore-producing bacteria that can cause foodborne and other opportunistic infections (e.g., gas-gangrene) (Kierzkowska *et al.*2018). In addition, *Corynebacteriaceae* and *Enterococcaceae*were found in relatively high abundance in the insects' microbiota. The family *Corynebacteriaceae* includes species of the genus*Corynebacterium* that cause opportunistic infections in humans, such as diphtheria, endocarditis, and lymphadenitis (Zhi *et al.*2017) . *Enterobacteriaceae* are the agents of many community - and hospital-acquired infections (Perez 2018) , including urinary tract infections, surgical wounds infections and bacteraemia (Lebreton*et al.* 2014). Also found in the insect microbiota were*Neisseriaceae* . These bacteria can be a part of the normal human bacteria flora and are typically harmless. However, some species of this family can cause infections in humans (Osses *et al.* 2017); for example, *Neisseria gonorrhoeae*

and *Neisseria meningitidis*cause the sexually transmitted disease gonorrhoea and meningitis, respectively (Christodoulides *et al.* 2021).

Less abundant bacterial families in the insect captures included*Peptostreptococcaceae* , commonly found in the gut of humans and animals (Neumann *et al.* 2020); the opportunistic bacteria family*Pseudomonadaceae* associated with antibacterial resistance (Saati-Santamaria *et al.* 2021; Wessels *et al.* 2021); and*Staphylococcaceae* . The latter are common human commensals and typically harmless. However, they can cause infections in immunocompromised people (Rossi *et al.* 2020).*Streptococcaceae* are associated with infective endocarditis, purulent infections, brain haemorrhage, intestinal inflammation and bacteraemia (Yumoto *et al.* 2019). The identification of human pathogens that can cause opportunistic infections in the indoor environment does not directly translate into a public health concern. However, our findings invite a closer investigation of the microbiota of non-pest arthropods as a potential reservoir of opportunistic bacteria that can cause pathogenic effects in humans. With regular screening of household environments, opportunistic infections may be prevented.

Origin of the arthropod microbiota

Arthropod endogenous microbiota is typically adapted to the host and genetically determined (Smee *et al.* 2021; Suppa *et al.*2020). However, environmentally acquired bacteria are also found in both endogenous and exogenous microbiota of arthropods. These bacteria may have specific regulatory responses and, therefore, may be selectively acquired from the environment in each generation (Engel & Moran 2013; Hannula *et al.* 2019). Typically, the exogenous microbiota is more diverse than the endogenous one (Oliveira Ramalho *et al.* 2019).

Consistent with previous studies, we found that the captured arthropods had significantly distinct microbiota and that the diversity of these microbial communities varied among morphospecies showing, on average, a lower diversity in endogenous than exogenous microbiota. Arthropods vary in the extent to which microorganisms are essential to their survival and fitness, with extremes represented by insects with little to no gut microbiota but dependent on intracellular symbionts (e.g., sap feeding insects), to insects with large and complex communities (e.g., termites). Most insects fall within this range with most having moderately complex gut microbiota with ca. 20 abundant taxa (Hammer*et al.* 2017; McFall-Ngai 2007). Less is known about the diversity of the exogenous microbiota associated with arthropods.

We found that human pathogenic bacteria were abundant in both the endogenous and exogenous microbiota of the studied arthropods. However, these pathogens were only present in indoor arthropods ecologically associated with humans (synanthropic arthropods), and absent in the microbiota of outdoor arthropods. This suggests that the microbiota may be influenced by the environment, but that the establishment of environmental microbes is genetically determined.

The culture-based approach showed severe limitations as compared to the unbiased metabarcoding approach. Through metabarcoding, we were able to identify a higher number of bacteria families and to obtain an unbiased characterization of the arthropod's associated microbiota. The metabarcoding approach was also able to identify a higher number of potentially opportunistic pathogens associated with the arthropods, improving our understanding of the role of non-pest arthropods as carriers of human pathogens. A parallel analysis of opportunistic infections in humans and the opportunistic pathogens in indoor the arthropods is necessary to establish the risk associated with non-pest arthropods as carriers of human diseases. However, our study provides important insights into the role that non-pest arthropods may have as carriers of opportunistic pathogens in households.

Acknowledgements

This research was funded by F. Boiocchi's PhD scholarship (RMB/ACH/TW/MD/JT 50th Anniv. 'The authors are thankful to Stephen Kissane for the preparation and sequencing of the metabarcoding libraries and to the citizen scientists.

Author contributions

FB conducted the experiments, analyzed the data and drafted the manuscript; RD contributed to data analysis of high throughput sequencing data; MPD assisted with the arthropods identification; LO and ACH designed the experiment, guided the data analysis process and finalized the manuscript. All authors contributed to the manuscript writing.

Data availability

DNA sequences obtained with Sanger technology have the following accession numbers: MZ700565 - MZ700685. The metabarcoding sequences can be found at BioProject PRJNA762094 in the NCBI repository.

References

(ACDP) ACoDP (2020) Health and Safety Executive Annual Report and Accounts. Public Health England.

Adams RI, Bateman AC, Bik HM, Meadow JF (2015) Microbiota of the indoor environment: a meta-analysis. *Microbiome* **3** , 49.

Barberan A, Dunn RR, Reich BJ*, et al.* (2015) The ecology of microscopic life in household dust. *Proc Biol Sci* **282** .

Barreiro C, Albano H, Silva J, Teixeira P (2013) Role of flies as vectors of foodborne pathogens in rural areas. *ISRN Microbiol***2013** , 718780.

Bertone MA, Leong M, Bayless KM*, et al.* (2016) Arthropods of the great indoors: characterizing diversity inside urban and suburban homes.*PeerJ* **4** , e1582.

Brown PMJ, Roy HE, Rothery P*, et al.* (2008) Harmonia axyridis in Great Britain: Analysis of the spread and distribution of a non-native coccinellid. *BioControl* **53** , 55–67.

Callahan BJ, McMurdie PJ, Rosen MJ*, et al.* (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13** , 581-583.

Caporaso JG, Kuczynski J, Stombaugh J*, et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7** , 335-336.

Chinery M (1993) Insects of Britain and Northern Europe. H. London,Ed.

Chinery M, Falk SJ (2007) Domino guide to the insects of Britain and Western Europe. In: *London, UK* . A&C Black Publishers Ltd.

Christodoulides M, Humbert MV, Heckels JE (2021) The potential utility of liposomes for Neisseria vaccines. *Expert Rev Vaccines***20** , 1235-1256.

Das JK, Hadi YB, Salam RA*, et al.* (2018) Fly control to prevent diarrhoea in children. *Cochrane Database Syst Rev* **12** , CD011654.

Desales-Lara MA, Francke OF, Sánchez-Nava P (2013) Diversidad de arañas (Arachnida: Araneae) en hábitats antropogénicos. *Revista Mexicana de Biodiversidad* **84** , 291–305.

Doron S, Gorbach S (2008) Bacterial Infections: Overview. In:*International Encyclopedia of Public Health* , pp. 273–282.

Durán-Barrón CG, Francke OF, Pérez-Ortiz TM (2009) Diversidad de arañas (Arachnida: Araneae) asociadas con viviendas de la ciudad de México (Zona Metropolitana). *Revista Mexicana de Biodiversidad***80** , 55–69.

Engel P, Moran NA (2013) The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol Rev* **37** , 699-735.

Faulde M, Spiesberger M (2013) Role of the moth fly Clogmia albipunctata (Diptera: Psychodinae) as a mechanical vector of bacterial pathogens in German hospitals. *J Hosp Infect* **83** , 51-60.

Gibbons SM (2016) The Built Environment Is a Microbial Wasteland.*MSystems* **1** , 1-4.

Hall RD, Gerhardt RR (2002) Flies (Diptera). In: *Medical and Veterinary Entomology* , pp. 127–145. Elsevier Science, USA.

Hammer TJ, Janzen DH, Hallwachs W, Jaffe SP, Fierer N (2017) Caterpillars lack a resident gut microbiome. *Proc Natl Acad Sci U S A* **114** , 9641-9646.

Hannula SE, Zhu F, Heinen R, Bezemer TM (2019) Foliar-feeding insects acquire microbiomes from the soil rather than the host plant. *Nat Commun* **10** , 1254.

Holland D, Thomson L, Mahmoudzadeh N, Khaled A (2020) Estimating deaths from foodborne disease in the UK for 11 key pathogens. *BMJ Open Gastroenterol* **7** .

Jeon YS, Chun J, Kim BS (2013) Identification of household bacterial community and analysis of species shared with human microbiome.*Curr Microbiol* **67** , 557-563.

Kierzkowska M, Pedzisz P, Babiak I*, et al.* (2018) Difficulties in identifying the bacterial species from the genus Clostridium in a case of injury-related osteitis. *Folia Microbiol (Praha)* **63** , 533-536.

Knight SM, Toodayan W, Caique WC*, et al.* (1992) Risk factors for the transmission of diarrhoea in children: a case-control study in rural Malaysia. *Int J Epidemiol* **21** , 812-818.

Krawczak FS, Labruna MB, Hecht JA, Paddock CD, Karpathy SE (2018) Genotypic Characterization of Rickettsia bellii Reveals Distinct Lineages in the United States and South America. *Biomed Res Int***2018** , 8505483.

Kwon HW, Yun YH, Kim JY*, et al.* (2016) Investigation of bacterial diversity in sciaridae flies collected in greenhouses used for shiitake cultivation. *Journal of Odor and Indoor Environment* **15** , 147–153.

Lebreton F, Willems RJL, Gilmore MS (2014) Enterococcus Diversity, Origins in Nature, and Gut Colonization. In: *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection* (eds. Gilmore MS, Clewell DB, Ike Y, Shankar N), Boston.

Leong M, Bertone MA, Savage AM*, et al.* (2017) The Habitats Humans Provide: Factors affecting the diversity and composition of arthropods in houses. *Sci Rep* **7** , 15347.

MacConaill LE, Burns RT, Nag A*, et al.* (2018) Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* **19** , 30.

Martin M (2011) Cutadapt removed adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17** , 10.

McFall-Ngai M (2007) Adaptive immunity: care for the community.*Nature* **445** , 153.

McKnight DT, Huerlimann R, Bower DS*, et al.* (2019) microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environmental DNA***1** , 14–25.

Memona H, Manzoor F, Anjum AA (2017) Cockroaches (Blattodea: Blattidae): A Reservoir of Pathogenic Microbes in Human-Dwelling Localities in Lahore. *J Med Entomol* **54** , 435-440.

Menasria T, Tine S, Mahcene D*, et al.* (2015) External Bacterial Flora and Antimicrobial Susceptibility Patterns of Staphylococcus spp. and Pseudomonas spp. Isolated from Two Household Cockroaches, Blattella germanica and Blatta orientalis. *Biomed Environ Sci* **28** , 316-320.

Meola M, Rifa E, Shani N*, et al.* (2019) DAIRYdb: a manually curated reference database for improved taxonomy annotation of 16S rRNA gene sequences from dairy products. *BMC Genomics* **20** , 560.

Moges F, Eshetie S, Endris M*, et al.* (2016) Cockroaches as a Source of High Bacterial Pathogens with Multidrug Resistant Strains in Gondar Town, Ethiopia. *Biomed Res Int* **2016** , 2825056.

Neumann A, Bjorck L, Frick IM (2020) Finegoldia magna, an Anaerobic Gram-Positive Bacterium of the Normal Human Microbiota, Induces Inflammation by Activating Neutrophils. *Front Microbiol***11** , 65.

Oberauner L, Zachow C, Lackner S*, et al.* (2013) The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. *Sci Rep* **3** , 1413.

Ogata H, La Scola B, Audic S*, et al.* (2006) Genome sequence of Rickettsia bellii illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet* **2** , e76.

Oksanen J (2020) Vegan: an introduction to ordination, pp. 1-10.

Oliveira Ramalho M, Saux Moreau C, Correa Bueno O (2019) The Potential Role of Environment in Structuring the Microbiota of Camponotus across Parts of the Body. *Advances in Entomology,* **7** , 47-70.

Osses DF, Dijkmans AC, van Meurs AH, Froeling FM (2017) Neisseria Mucosa: A New Urinary Tract Pathogen? *Curr Urol* **10** , 108-110.

Perez LRR (2018) Nosocomial impact of prevalent beta-lactamases from the community enterobacteriaceae: what to do when the resistance doesn't go your way. *Infect Control Hosp Epidemiol* **39** , 1016-1017.

Rahmadani F, Lee H (2020) Dynamic Model for the Epidemiology of Diarrhea and Simulation Considering Multiple Disease Carriers. *Int J Environ Res Public Health* **17** .

Ranjbar R, Izadi M, Hafshejani TT, Khamesipour F (2016) Molecular detection and antimicrobial resistance of Klebsiella pneumoniae from house flies (Musca domestica) in kitchens, farms, hospitals and slaughterhouses. *J Infect Public Health* **9** , 499-505.

Reed HC, Landolt PJ (2019) Ants, wasps, and bees (Hymenoptera). In:*Medical and Veterinary Entomology* , pp. 459-488. Academic Press.

Rodriguez-Rodriguez SE, Solis-Catalan KP, Valdez-Mondragon A (2015) Diversity and seasonal abundance of anthropogenic spiders (Arachnida: Araneae) in different urban zones of the city of Chilpancingo, Guerrero, Mexico. *Revista Mexicana de Biodiversidad* **86** , 962–971.

Rossi CC, Pereira MF, Giambiagi-deMarval M (2020) Underrated Staphylococcus species and their role in antimicrobial resistance spreading. *Genet Mol Biol* **43** , e20190065.

Saati-Santamaria Z, Peral-Aranega E, Velazquez E, Rivas R, Garcia-Fraile P (2021) Phylogenomic Analyses of the Genus Pseudomonas Lead to the Rearrangement of Several Species and the Definition of New Genera.*Biology (Basel)* **10** .

Sengupta PG, Sircar BK, Mandal SK*, et al.* (1995) Epidemiology of Vibrio cholerae O139 with special reference to intrafamilial transmission in Calcutta. *J Infect* **31** , 45-47.

Smee MR, Raines SA, Ferrari J (2021) Genetic identity and genotype x genotype interactions between symbionts outweigh species level effects in an insect microbiome. *ISME J* **15** , 2537-2546.

Stenfors Arnesen LP, Fagerlund A, Granum PE (2008) From soil to gut: Bacillus cereus and its food poisoning toxins. *FEMS Microbiol Rev***32** , 579-606.

Suppa A, Kvist J, Li X*, et al.* (2020) Roundup causes embryonic development failure and alters metabolic pathways and gut microbiota functionality in non-target species. *Microbiome* **8** , 170.

Team RC (2020) R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing* , Vienna, Austria.

Voloshyn V, Tymchuk K, Symochko L, Kacaniova M, Fedoriak M (2017) Spiders and Other Arthropods of Chernivtsi Poultry Farm (Ukraine) and the Preliminary Data About Bacteria Inhabiting Their External Surfaces.*International Journal of Ecosystems and Ecology Science-Ijees***7** , 587–596.

Wang S, Michaud J, Tan X, Zhang F, XJ G (2011) The Aggregation Behavior of Harmonia axyridis in its Native Range in Northeast China.*BioControl* **56** , 193–206.

Wessels L, Reich F, Schmoger S*, et al.* (2021) Draft Genome Sequences of Pseudomonas sp. Isolates Recovered from Ghanaian Fish Food Samples in 2018. *Microbiol Resour Announc* **10** .

Yumoto H, Hirota K, Hirao K*, et al.* (2019) The Pathogenic Factors from Oral Streptococci for Systemic Diseases. *Int J Mol Sci***20** .

Zhi XY, Jiang Z, Yang LL, Huang Y (2017) The underlying mechanisms of genetic innovation and speciation in the family Corynebacteriaceae: A phylogenomics approach. *Mol Phylogenet Evol* **107** , 246-255.

Data Accessibility

DNA sequences obtained with Sanger technology have the following accession numbers: MZ700565 - MZ700685. The metabarcoding sequences can be found at BioProject*PRJNA762094*in the NCBI repository.

Author ContributionsFB conducted the experiments, analyzed the data and drafted the manuscript; RD contributed to data analysis of high throughput sequencing data; MPD assisted with the arthropods identification; LO and ACH designed the experiment, guided the data analysis process and finalized the manuscript. All authors contributed to the manuscript writing.

Table 1 PERMANOVA . Permutational Multivariate Analysis of Variance testing for the effect of Arthropod Family, Environment (urban/suburban), sampling month (Month) and their interaction terms in the culture-based (A) and metabarcoding experiment (B). The Bray-Curtis distance (Oksanen 2020) was used in both analyses. Significant terms are in bold. The statistics in this Table support data in Figure 3.

	Df	**F.Model**	**R2**	**Pr(>F)**
Culture-based	Culture-based	Culture-based	Culture-based	Culture-based
Arthropod Family	8	1.88	0.16	**0.001**
Environment (Env)	1	2.15	0.02	**0.033**
Season	11	1.15	0.13	0.126
Arthropod Family*Env	3	1.22	0.04	0.16
Arthropod Family*Season	6	1.17	0.07	0.162
Season*Env	5	1.08	0.05	0.283
Metabarcoding	Metabarcoding	Metabarcoding	Metabarcoding	Metabarcoding
Arthropod Family	8	1.72	0.086	**0.005**
Environment (Env)	1	1.89	0.01	**0.021**
Season	11	2.55	0.13	**0.002**
Arthropod Family*Env	1	0.57	0.003	0.8
Arthropod Family*Season	13	1.37	0.111	**0.01**
Season*Env	2	1.7	0.02	0.073

Table 2 PERMANOVA. Permutational Multivariate Analysis of Variance testing for the effect of Arthropod Family, Environment (urban/suburban), sampling month (Month) and their interaction terms on exogenous (exo) and endogenous (endo) microbiota in the culture-based (A) and metabarcoding experiment (B). Bray-Curtis (Oksanen 2020) distance was used on both culture-based approach and on ASV matrices. Significant terms are in bold.

	Df	**F.Model**	**R2**	**Pr(>F)**
Culture-based	Culture-based	Culture-based	Culture-based	Culture-based
Arthropod Family	4	2.3	0.17	**0.002**
Exo/endo	1	0.44	0.008	0.858

Season	2	0.78	0.02	0.663
Family*exo/endo	4	2.36	0.17	**0.001**
Exo/endo*Season	1	0.44	0.08	0.845
Metabarcoding	Metabarcoding	Metabarcoding	Metabarcoding	Metabarcoding
Arthropod Family	4	10.01	0.38	**0.001**
Exo/endo	1	7.28	0.07	**0.001**
Season	3	1.52	0.04	0.077
Family*Exo/endo	4	3.61	0.14	**0.001**
Exo/endo*Season	3	0.93	0.02	0.5

Figure Legends

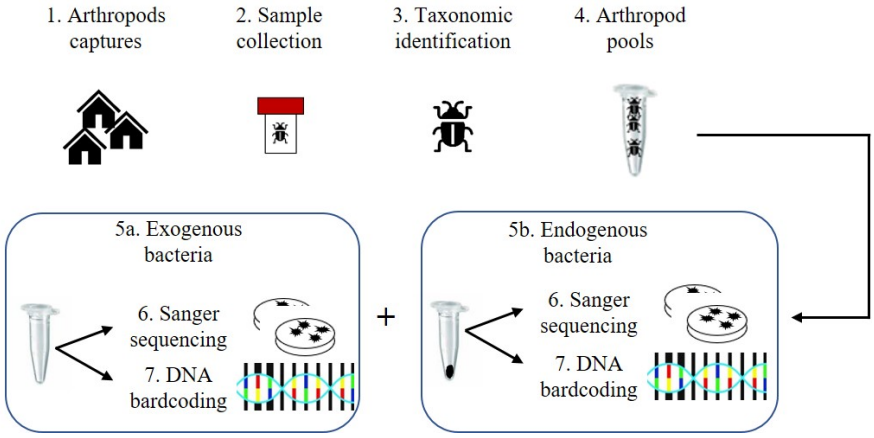
Figure 1 Experimental design . Arthropods were actively captured by citizen scientists from 20 households between August 2018 and October 2019 (1); the samples were collected monthly by researchers at Aston University, UK (2); the arthropods were taxonomically classified (3) and pools of up to 5 specimens of the same morphospecies pooled (4). From these pools endogenous and exogenous microbiota were isolated (5); aliquots of these isolates were plated on different substrates and sequenced using Sanger sequencing (6) and sequenced using DNA barcoding with the V1-V2 region of the 16S (7).

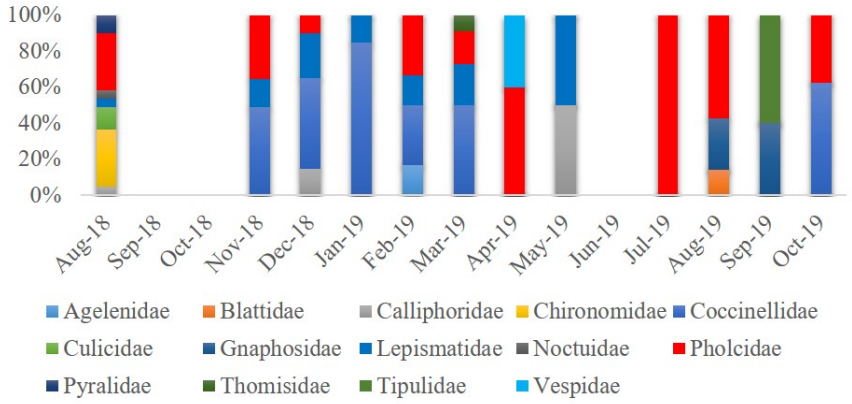
Figure 2. Arthropods relative abundance. Stacked barplot of arthropods' relative abundance between August 2018 and October 2019. Barplots are blank when data were not collected by the citizen scientists.

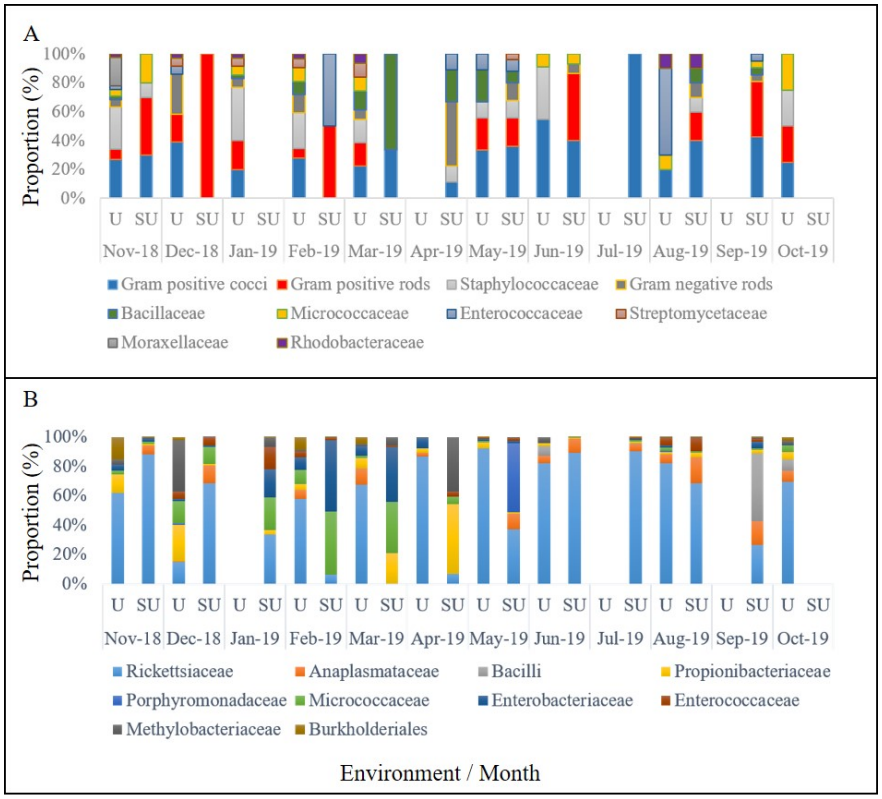
Figure 3. Microbiome diversity with culture-based and metabarcoding approaches. The relative abundance of the top 10 most abundant bacterial families diversity is shown in stacked barplots for A) the culture-based and B) the metabarcoding approach. Data are shown for urban (U) and suburban (SU) households per month of sampling. Statistical analyses supporting this Figure are in Table 1.

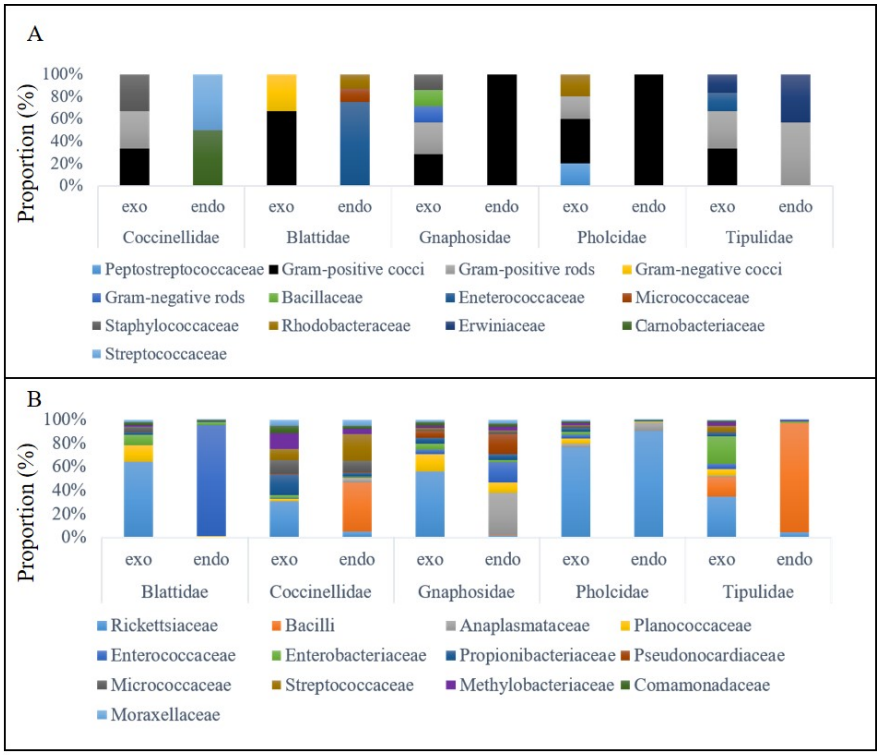
Figure 4. Exogenous and endogenous microbiota composition. Stacked bar plots representing the microbiota composition of exogenous (exo) and endogenous (endo) bacterial communities, across the sampling months for the culture-based (A) and the metabarcoding approach (B). No culturable bacteria were isolated from the endogenous microbiota in July 2019 with the culture-based approach. Statistical analyses supporting this Figure are in Table 2.

Figure 5. Shared bacteria families . Venn diagram showing the intersection among bacteria families identified by the culture-based and the metabarcoding approach, and their overlap with the approved list of biological agents by the Advisory Committee on Dangerous Pathogens (HSE).









Month/microbiota